Review article

Michael Denker, Sonja Grün, Thomas Wachtler and Hansjörg Scherberger*

Reproducibility and efficiency in handling complex neurophysiological data

https://doi.org/10.1515/nf-2020-0041

Abstract: Preparing a neurophysiological data set with the aim of sharing and publishing is hard. Many of the available tools and services to provide a smooth workflow for data publication are still in their maturing stages and not well integrated. Also, best practices and concrete examples of how to create a rigorous and complete package of an electrophysiology experiment are still lacking. Given the heterogeneity of the field, such unifying guidelines and processes can only be formulated together as a community effort. One of the goals of the NFDI-Neuro consortium initiative is to build such a community for systems and behavioral neuroscience. NFDI-Neuro aims to address the needs of the community to make data management easier and to tackle these challenges in collaboration with various international initiatives (e.g., INCF, EBRAINS). This will give scientists the opportunity to spend more time analyzing the wealth of electrophysiological data they leverage, rather than dealing with data formats and data integrity.

Keywords: FAIR; NFDI; open data; research data management; systems neuroscience.

Zusammenfassung: Die Aufbereitung eines neurophysiologischen Datensatzes mit dem Ziel, ihn zu teilen und zu veröffentlichen, ist schwierig. Viele der verfügbaren Werkzeuge und Dienste für einen reibungslosen Ablauf einer Datenpublikation sind noch im Entstehen und nicht gut integriert. Außerdem fehlen Handlungsempfehlungen und konkrete Beispiele für die Publikation eines vollständigen Datensatzes aus elektrophysiologischen Experimenten. Angesichts der Heterogenität des Feldes können solche einheitlichen Richtlinien und Prozesse nur gemeinschaftlich formuliert werden. Eines der Ziele der Konsortiumsinitiative NFDI-Neuro ist es, für die System- und Verhaltensneurowissenschaften eine solche Gemeinschaft aufzubauen. NFDI-Neuro will die Bedürfnisse dieser Community für ein verbessertes Datenmanagement aufgreifen und in Zusammenarbeit mit verschiedenen internationalen Initiativen (z.B. INCF, EBRAINS) angehen und lösen. Hierdurch bleibt den Wissenschaftlern in Zukunft mehr Zeit zur Analyze ihrer reichhaltigen elektrophysiologischen Daten, anstatt sich mit Datenformaten und Datenintegrität befassen zu müssen.

Schlüsselwörter: FAIR; NFDI; offene Daten; Forschungsdatenmanagement; Systemneurowissenschaften.

***Corresponding author: Hansjörg Scherberger**, Neurobiology Laboratory, Deutsches Primatenzentrum GmbH, Kellnerweg 4, 37077 Göttingen, Germany; and Department of Biology and Psychology, University of Goettingen, Goettingen, Germany, E-mail: hscherb@gwdg.de. https://orcid.org/0000-0001-6593-2800
Michael Denker, Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany, E-mail: m.denker@fz-juelich.de. https://orcid.org/0000-0003-1255-7300
Sonja Grün, Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany, E-mail: s.gruen@fz-juelich.de. https://orcid.org/0000-0003-2829-2220
Thomas Wachtler, Department Biologie II, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany, E-mail: wachtler@bio.lmu.de. https://orcid.org/0000-0003-2015-6590

Introduction

Neurophysiology may be considered one of the most common approaches in neuroscience to gain an understanding of the internal processes that underlie neuronal information processing by examining neural activity at various scales of observation. The technique profits from constant technological evolution of recording techniques that enable ever more intricate experimental designs to investigate neuroscientific questions. Although our growing insight into neuronal computation is manifested in the development of increasingly realistic models of brain dynamics through simulation and theory, electrophysiology arguably continues to provide the most valuable experimental counterpart upon which the process of cross-validation is based.

However, the new opportunities offered by the rapid technological and conceptual developments during the last decades do not come for free. The increase in complexity of modern-day experimentation is mirrored in an intricate pool of data, resulting from various hardware and software components built by industrial manufacturers or in-house workshops, and combined in versatile ways to enable novel experimental designs. The amount of experimental skill, time, and creativity that enters such experiments leads to a situation where the protocols conducted in individual labs are to a large extent unique to the respective research group. In order to describe such experiments at a level of being reproducible (Denker and Grün, 2016; Plesser, 2018) and the resulting data becoming practically findable and reusable by other researchers, as stipulated by the FAIR principles (Wilkinson et al., 2016; see also Wachtler et al., this issue), there is a substantial conceptual difficulty in documenting data acquisition and postprocessing at minute levels of detail due the inherent heterogeneity and complexity. Moreover, in the absence of automatization, as typically associated with more standardized processes in science, the costs for a thorough description of the recorded data seem prohibitive. Therefore, the intricate nature of this heterogeneous data bundle, together with the need to integrate all data into a form that is suitable for the anticipated analysis, leads to situations where scientists come up with ad-hoc and highly customized data analysis solutions. This is not only time consuming and inefficient but also error prone and hardly reproducible.

The increased complexity of experimentation is a challenge for organizing the data, and it also amplifies the urgency to conceptualize and implement better solutions of data management. Firstly, while data from electrophysiology have always been considered particularly precious, given the amount of invested resources (time, money, and animal life), the richness and complexity of modern experiments further increase their value by expanding the number of possible scientific questions that can be addressed by a single data set. Indeed, the primary use of the original experiment often leverages only a small part of the full potential of the data. In consequence, data sets may be of relevance for research questions a long time after the original recording and for a large audience, which makes it infeasible for the original experimenters to guide and check that data are handled appropriately. Such scenarios include the notorious case of the PhD student leaving a research group without providing sufficient information about the recorded data and the preprocessing and postprocessing steps applied to them during the PhD project. Another case is the situation where data are analyzed in parallel by a number of laboratories applying different methods and having different research questions with the aim to synergize their findings, like in a collaboration between an experimental and a theoretical lab. For the exchange of results and findings, these partners need to have a consistent description of the data. Even beyond the initial use of a data set within a laboratory and among their collaborators, more and more scientists embrace the idea of making their data publically available, recognizing not only the added attention and appreciation that well-curated data are generating for the experimentalist, but also the increased efficiency for progressing science and the potential to create new research questions.

Despite all of these suggested merits of striving to handle and manage electrophysiology data in a more optimal fashion, we perceive that reality is far from the ideal situation where all of the essential, cumbersome housekeeping of acquired data is automated and the description of the experiment can be saved with the proverbial click of a button. Yet, the topic of data handling is currently erupting in a burst of activity in the field of (neuro-)informatics. Nevertheless, we still observe a prevailing gap between the design of emerging tools, services and processes, and their implementation in concrete experimental settings that are helpful for experimentalists.

The anticipated NFDI-Neuro consortium therefore considers its task to communicate between the world of computer science and neuroscience and to enable the neuroscience community to make better use of the existing tools and services. On the other hand, NFDI-Neuro also supports the neuroscience community to link its established data acquisition and postprocessing workflows to existing tools and to identify lacking tools, processes, and guidelines. The heterogeneity of the data and experimental approaches in the neuroscience community demands to base discussions on concrete examples and build on experiences. NFDI-Neuro considers the establishment of such an exchange to be a primary goal for its activities surrounding electrophysiology.

To stimulate such a discussion, we report here on current challenges, solutions, and shortcomings as we encountered them in our research routines (e.g., Zehl et al., 2016) and during efforts to publish an experimental data set consisting of spiking activity and local

field potentials of a macaque monkey performing a motor task (Brochier et al., 2018). We will outline the experiences we had in curating such a data set, ranging from dealing with a collaborative environment consisting of multiple labs, up to an ongoing and dynamic data acquisition process established over the course of years, and the need to publish and easily maintain data in an accessible format.

Considerations when curating electrophysiological data sets

Adoption of any type of data management workflow should ideally disrupt the research workflow only minimally and require only little attention of the researcher. One of the most rewarding strategies is therefore to select data curation procedures that remain constant, consistent, and mostly automated. Given the unpredictable nature of the research process, it may, however, be tempting to design the data curation process "on the fly" as the experiment is being set up. However, anticipating the future use of the data can help guide design decisions early on, which expedite the establishment of a stable data curation pipeline. This includes the consideration of the perhaps most challenging scenario from the start: sharing data with strangers. By design, this approach will help to perceive the process of data acquisition from the perspective of a data consumer (e.g., the remotely analyzing scientist) not the data provider (i.e., the experimenter). Such considerations will address a diversity of issues such as making the data and corresponding metadata available in formats independent of specific programming languages, avoidance of idiosyncratic software codes, and favoring easily comprehensible data descriptions over those present in the original hardware and control software implementations.

For example, if data are provided as "raw" (or primary) data, representing directly the output of the recording setup, users accessing these data need to understand the specific data structure or, equivalently, require specific software for reading the data. While this is conceptually a feasible approach, in practice, it may often fail even at the level of reading the raw binary data since codes for reading the corresponding file formats often differ in the various programming languages and rarely receive professional maintenance and thorough testing. When it comes to interpreting the data contained in these raw files, matters tend to become even more difficult. For example, in typical recordings, raw data contain only certain marked events in time, such as events indicating the start of a certain experimental trial or the time point of a stimulus presentation. Uniquely identifying and describing trials, however, is a long way from these marked events. For example, it may involve the need to interpret the type of a given trial based on subsequent events, in case alternative stimuli, manipulations, or behaviors are possible. It may also involve the interpretation of the performance in a trial based on separate behavioral measurements, e.g., reaction time. Ultimately, each trial must be labeled by an informative identifier that is based on this information and that supports the implementation of the planned analysis of the data. Finally, the outcome of these preparatory steps performed on the primary data (e.g., the resulting trial identification) needs to be completely consistent for any user of the data; otherwise, the comparison of the results of the different variations of data analysis is not reliable and will be reduced to an act of belief.

A second aspect to consider in developing the data curation workflow is that data are ideally made available for initial inspection soon after the first recordings are performed and should then already resemble the anticipated final output structure. Failure to analyze the data set early on bears the danger that potential shortcomings in the data are not noticed at an early stage. However, starting the data analysis using ad-hoc and makeshift solutions may prevent the later adoption of a more rigorous data management concept for those projects, for example, because they might not be not backward compatible. In such a situation, scientists who rely on such initial solutions might therefore not profit from future adjustments in data acquisition or postprocessing workflows, for example, to account for adaptations in postprocessing parameters or to incorporate additional descriptive metadata that were previously not considered.

Adopting the view of the naïve data consumer and the resulting need of a rigorous, comprehensible, and unambiguous data output, it became clear that both a defined process for data acquisition and postprocessing, as well as stable tools to implement corresponding standards in support of this process, are required (see Figure 1). The need for early access to the processed data, while allowing adjustments to the process as the experiment progresses, further indicated that going from the raw recorded data to the resulting processed data package must be reproducible

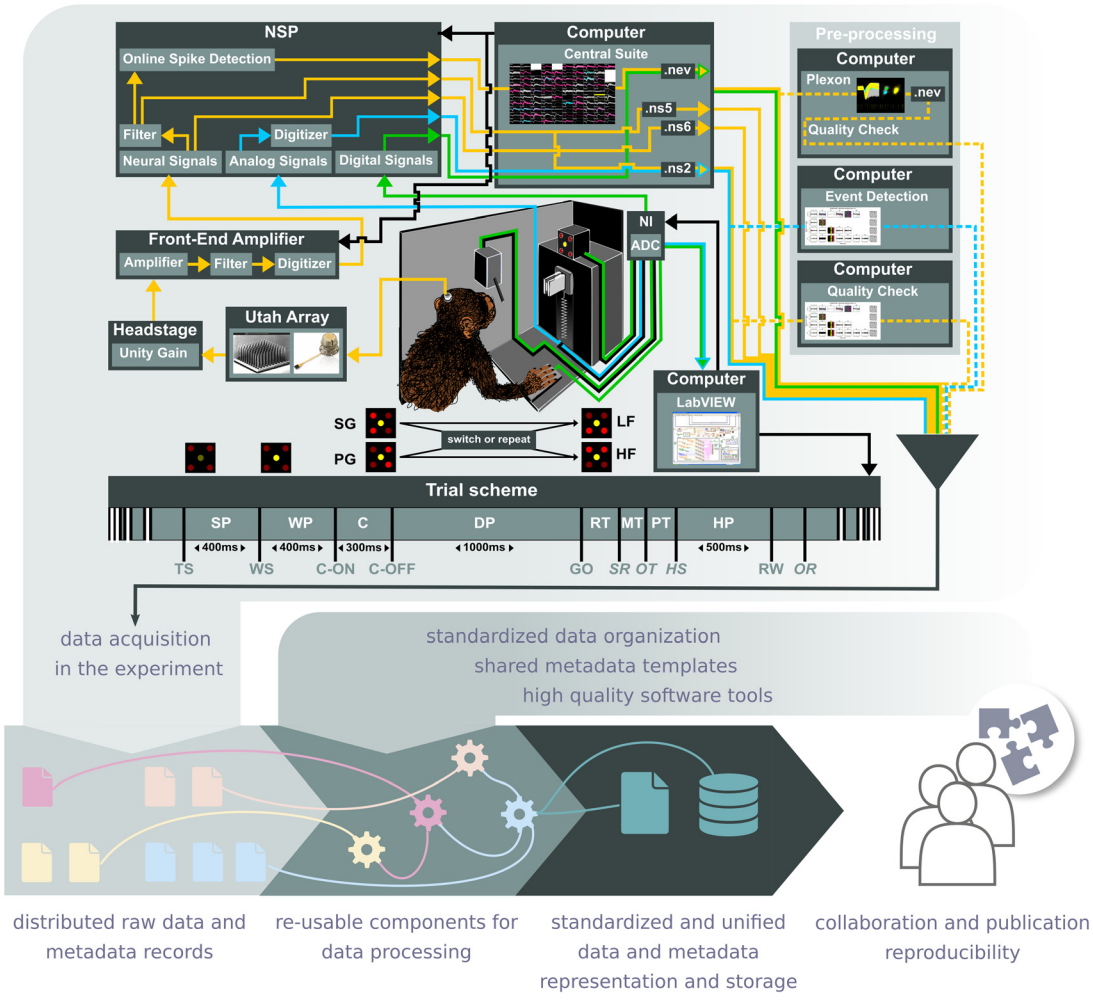

Figure 1: Data acquisition and postprocessing workflows simplify the process of preparing data for analysis, collaboration, and sharing. Data and metadata recorded in electrophysiological experiments are typically distributed across multiple disconnected files and stored in various proprietary and custom file formats. Development of standardized components for processing the recorded data based on common standards for data organization, shared ontologies and terminologies for metadata, and high-quality community tools leads to a comprehensive data representation that facilitates collaboration and fast publication via repositories. Top image modified from Zehl et al., 2016 (licensed under CC-BY).

at any time. As described in Zehl et al. (2016), such an approach requires – as much as possible – an automated way of building both the skeletal structure of data and metadata for the complete experiment and filling this structure with a particular data set. Such a generating process is able to produce a complete and consistent data package that is robust with respect to subsequent structural changes during the lifetime of the experiment, i.e., in case the data structure or specific parts of the metadata descriptions need to be adjusted.

While the implementation of such a process for data curation ultimately leads to a well-documented data set, designing this workflow is laborious when performed from scratch, since it requires the researcher to consider the design decisions of the process in minute detail. Yet, while experiments differ, individual parts of this curation workflow can most likely be singled out and used for multiple experiments as easily adaptable building blocks. Sharing and reuse of such components of data curation workflows are therefore key elements to facilitate researchers to produce comparable, complete, and versatile workflows in reasonable time before starting the experiment and thus enable a more structured way to implement the data curation process. The work programme of NFDI-Neuro supports this process by establishing mechanisms for researchers to share their data acquisition, postprocessing

and analysis pipelines, to identify commonalities, and to produce common components for building individual data acquisition pipelines.

Metadata

When it comes to the underlying tools and services that enable such a data curation process, luckily the field is in a more advanced position. A first challenge we faced was the high degree of data fragmentation and descriptive metadata in the form of different files and file formats that need to be bundled for proper data curation. Here, the open metadata markup language (odML) offers an easy approach to adopt a machine readable and fully flexible data model that supports structuring and storing such metadata (Grewe et al., 2011). In this way, the wealth of details related to the experiment and each data set could be exposed to collaborators in an easily understandable manner. Still, the design of the actual metadata hierarchy for a particular experiment is a challenge, in particular in absence of standardized vocabularies or ontologies that suggest, based on prior experience, what metadata to record and how to label them.

More and more efforts are being drawn up to alleviate the problem of metadata organization and storage. These are in part proposed by data repository providers where metadata help in making data sets discoverable and interpretable, e.g., the CRCNS database (Teeters et al., 2008; http://crcns.org), EEGBase (Papez and Moucek, 2013), GIN (https://gin.g-node.org), or detailed metadata schemas that are developed as part of the EBRAINS curation service (https://ebrains.eu/service/share-data), such as openMINDS to describe high-level minimal metadata (cf., https://github.com/HumanBrainProject/openMINDS). Also, for more in-depth metadata describing further experimental details, efforts have started to pool and harmonize metadata templates for different experimental aspects like hardware components, experimental paradigms, and measurement techniques (cf., e.g., Bower, 2009). The emerging terminologies are commonly based on community contributions and published work, such as the G-Node terminologies (https://terminologies.g-node.org/), NeuroElectro (Tripathy et al., 2014; https://neuroelectro.org/), ontologies, and terminologies provided as part of the NIF information framework (Imam et al., 2012; https://neuinfo.org/). In part, metadata schemas inspired by computational neuroscience are equally relevant for neurophysiology, e.g., NeuronDB and ModelDB (Hines et al., 2004; https://senselab.med.yale.edu/neurondb).

One of the main goals of NFDI-Neuro's task area for electrophysiological data will be to work toward making these resources interoperable and easy to integrate into a detailed data acquisition and postprocessing workflow already at the planning stage of the experiment. For this, we envision that components of the workflow provide automatic metadata for stereotypical processing steps and assist in finding appropriate metadata descriptions for those parts of the curation workflow that require customization with respect to the specific experiment.

Data formats

The next question we encountered was that of choosing a data format in which the final data packet would be available. Typically, recorded data are stored in files, often using a file format specified by the manufacturer of the recording system. The researcher is then presented with two possible scenarios:

(1) The shared data files are left untouched and are accompanied by a piece of code that loads these data and metadata in accordance with the experiment.
(2) Alternatively, a new data file is created that contains the annotated and curated data and metadata in a standardized format.

Either option has advantages and disadvantages. In the first scenario, data duplication is minimal, an important factor for experiments generating large quantities of data. Moreover, keeping the original data minimizes the risk of potential errors in moving data from one representation to another. On the downside, the recipient of the shared data will be presented with a proprietary data format that requires a highly customized loading routine. Such loading routines are in danger of becoming outdated over time and rarely receive testing by a larger community to prevent errors. In the second scenario, this danger can be prevented by supplying the data set in a standardized format that is read by well-tested and maintained loading code that tends to be more stable over time.

More importantly, the second scenario has two further advantages. First, a common standard data file format will simplify the use of curated data in multiple programming languages. For example, at the time of publication of our data set (Brochier et al., 2018), these file formats were not yet sufficiently mature; consequently, a second set of data files (mat-format for Matlab users) had to be supplied in addition to the original data files and the Python code,

thus duplicating storage space. Second, many vendor-specific formats are designed from the perspective of the recording system, i.e., data packets from multiple channels and are written progressively to file. From a consumer perspective, however, one of the most common scenarios is to read recording traces of one or several selected channels. The corresponding data samples are distributed across the raw data file, causing suboptimal performance in loading and processing. In contrast, standardized data formats provide more efficient data and metadata storage for the end user.

Both scenarios require a well-defined access to data stored in the various file formats. Efforts to form an alliance with manufacturers to provide a common, platform-independent, and well-tested basis for data access is still far from reality, despite early efforts by the Neuroshare initiative (http://neuroshare.sourceforge.net/index.shtml) aiming to unify access to various file formats from different vendors. In the Python world, the Neo data object model (Garcia et al., 2014) currently hosts the most comprehensive set of loading routines resulting from a community effort, which anticipates synergy with the SpikeInterface project aimed at evaluating spike sorter performance (Buccino et al., 2020; https://spikeinterface.readthedocs.io). In this design concept, data are represented in a common, generic structure independent of the source, which provides easy data access in a generic fashion from applications, analysis scripts, or other components of the data processing workflow.

When pursuing the first scenario of data publication (as we did for the data set described in Brochier et al., 2018), the most efficient and robust approach to construct the accompanying code for data access was to rely on a public community library such as Neo to handle the actual data loading and then annotate and reshape the resulting data object in a second step to optimally present its structure to the user. For the second scenario, data can be saved in a common, generic file format. Perhaps due to the high diversity of vendors of electrophysiological recording systems, such a common file format had not been available. However, to close this gap, two promising and complementary efforts have recently started. The first, Neurodata Without Borders (Teeters et al., 2015; https://www.nwb.org), offers a highly structured, HDF5-based format (Hierarchical Data Format version 5) to hold neurophysiological data sets based on a defined, optimized scheme. The second, NIX (Stoewer et al., 2014; http://www.g-node.org/nix), is a file format more customizable and suitable to combine structured data and arbitrary metadata records and fully compatible with odML-based metadata descriptions. Support to connect the Neo object model is continuously improved for both formats. With respect to the organization of data files at the file system level, ongoing efforts exist to extend structures such as the BIDS schema (Gorgolewski et al., 2016) to electrophysiology (cf., e.g., Pernet et al., 2019 for EEG, or discussions of the newly formed INCF special interest group on standardized data organization for electrophysiology), as are initiatives to establish interfaces with databases (e.g., Reimer et al., 2020).

NFDI-Neuro places a main focus on fostering these efforts toward common data models and file formats, and on making them interoperable with existing programming languages and storage solutions in the laboratories. Eventually, robust backing of data descriptions is key to ensure a smooth transition of electrophysiological data between any kind of data producer and data consumer.

Research data repositories

An important decision that must be made when deciding to share data is the physical storage location to use. Indeed, a number of data repositories exist to choose from, ranging from discipline-agnostic solutions such as Figshare (https://figshare.com) or Zenodo (https://zenodo.org), to generic institutional repositories, and to services catering specifically to the neuroscience community. Besides exposing the data set to a more targeted audience, the advantage when choosing one of the latter solutions is that these repositories are often able to interpret the contained data files as long as community standards are being adhered to. For example, the G-Node Infrastructure service used to store Brochier et al. (2018) is able to parse and display the odML encoded metadata schemes (see, e.g., https://gin.g-node.org/INT/multielectrode_grasp/src/master/datasets/i140703-001.odml), and the EBRAINS Knowledge Graph can link data sets to a corresponding view of its anatomical location in a brain atlas viewer.

Given the diversity of solutions that are available for sharing electrophysiological data, NFDI-Neuro's approach is to build a common infrastructure as a connecting layer, which will make access to data independent of specific storage solutions. In this way, researchers are able to choose the best repository for their data based on considerations of formal requirements, computational demands,

and capabilities of the repository store, while being able to simplify discoverability and access to the data.

Looking back, when we started to develop strategies to best share and publish data sets, we soon realized that a more efficient and robust process for data curation is essential in our field of science. However, it was only by way of experience that the prevailing gaps in our workflows became apparent. For this reason, we are confident that establishing a process within NFDI-Neuro to foster the interaction between the experimental realities in the laboratories and the development of sophisticated tools will lay the foundation that in the future scientists need to worry less about the technicalities of managing their data, but instead can appreciate the creativity sparked by analyzing the richness of state-of-the-art neural recordings.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This study is supported by LMUexcellent, Helmholtz Association, European Union's Horizon 2020 Framework Programme under grant no. 945539 (Human Brain Project SGA3), Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE), the German Federal Ministry of Education and Research (BMBF 01GQ1903), and the German Research Foundation (CRC889, FOR1847).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

Bower, M.R., Stead, M., Brinkmann, B.H., Dufendach, K., and Worrell, G.A. (2009). Metadata and annotations for multi-scale electrophysiological data. 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Minneapolis, MN: IEEE), pp. 2811–2814.

Brochier, T., Zehl, L., Hao, Y., Duret, M., Sprenger, J., Denker, M., Grün, S., and Riehle, A. (2018). Massively parallel recordings in macaque motor cortex during an instructed delayed reach-to-grasp task. Sci. Data *5*, 180055.

Buccino, A.P., Hurwitz, C.L., Garcia, S., Magland, J., Siegle, J.H., Hurwitz, R., and Hennig, M.H. (2020). SpikeInterface, a unified framework for spike sorting. eLife *9*, e61834.

Denker, M. and Grün, S. (2016). Designing workflows for the reproducible analysis of electrophysiological data. Brain-Inspired Computing. K. Amunts, L. Grandinetti, T. Lippert, and N. Petkov, eds. (Cham: Springer International Publishing), pp. 58–72.

Garcia, S., Guarino, D., Jaillet, F., Jennings, T., Pröpper, R., Rautenberg, P.L., Rodgers, C.C., Sobolev, A., Wachtler, T., Yger, P., et al. (2014). Neo: An object model for handling electrophysiology data in multiple formats. Front. Neuroinf. *8*, 10.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data *3*, 160044.

Grewe, J., Wachtler, T., and Benda, J. (2011). A bottom-up approach to data annotation in neurophysiology. Front. Neuroinf. *5*, 16.

Hines, M.L., Morse, T., Migliore, M., Carnevale, N.T., and Hines, M.L. (2004). ModelDB: A database to support computational neuroscience. J. Comput. Neurosci. *17*, 7–11.

Imam, F., Larson, S., Grethe, J., Gupta, A., Bandrowski, A., and Martone, M. (2012). Development and use of ontologies inside the neuroscience information framework: A practical approach. Front. Genet. *3*, 111.

Papez, V. and Moucek, R. (2013). Data and metadata models in electrophysiology domain: Separation of data models into semantic hierarchy and its integration into EEGBase. 2013 IEEE International Conference on Bioinformatics and Biomedicine (Shanghai, China: IEEE), pp. 539–543.

Pernet, C.R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., and Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. Sci. Data *6*, 103.

Plesser, H.E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. Front. Neuroinf. *11*, 76.

Reimer, M.L., Bangalore, L., Waxman, S.G., and Tan, A.M. (2020). Core principles for the implementation of the neurodata without borders data standard. J. Neurosci. Methods, 108972, https://doi.org/10.1016/j.jneumeth.2020.108972.

Stoewer, A., Kellner, C.J., Benda, J., Wachtler, T., and Grewe, J. (2014). File format and library for neuroscience data and metadata. Front. Neuroinform. Conference Abstract: Neuroinformatics 2014, https://doi.org/10.3389/conf.fninf.2014.18.00027.

Teeters, J.L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., et al. (2015). Neurodata without borders: Creating a common data format for neurophysiology. Neuron *88*, 629–634.

Teeters, J.L., Harris, K.D., Millman, K.J., Olshausen, B.A., Sommer, F.T. (2008). Data sharing for computational neuroscience. Neuroinformatics *6*, 47–55.

Tripathy, S.J., Savitskaya, J., Burton, S.D., Urban, N.N., and Gerkin, R.C. (2014). NeuroElectro: A window to the world's neuron electrophysiology data. Front. Neuroinf. *8*, 40.

Wachtler, T., Bauer, P., Denker, M., Grün, S., Hanke, M., Klein, J., Oeltze-Jafra, S., Ritter, P., Rotter, S., Scherberger, H., et al. (2021). NFDI-Neuro: Building a community for neuroscience research data management in Germany. Neuroforum, (this issue).

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018.

Zehl, L., Jaillet, F., Stoewer, A., Grewe, J., Sobolev, A., Wachtler, T., Brochier, T.G., Riehle, A., Denker, M., and Grün, S. (2016). Handling metadata in a neurophysiology laboratory. Front. Neuroinf. *10*, 26.

Bionotes

Michael Denker
Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany
m.denker@fz-juelich.de
https://orcid.org/0000-0003-1255-7300

Michael Denker received his diploma in physics from the University of Göttingen, Germany, in 2002. In 2004, he started his doctoral studies in the lab of Sonja Grün at the Free University Berlin. In 2006, he became a researcher at the RIKEN Brain Science Institute, Japan. He was awarded his PhD in 2009 at the Free University Berlin, Germany. In 2011, he joined the Institute of Neuroscience and Medicine (Research Center Jülich, Germany) and now leads the group Data Science in Electro- and Optophysiology Behavioral Neuroscience. His research interests are the analysis of the correlation structure of neural activity and its relationship to signals that express population activity and the establishment of workflows that improve the reproducibility of data analysis in neurophysiology.

Sonja Grün
Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany
Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany
s.gruen@fz-juelich.de
https://orcid.org/0000-0003-2829-2220

Sonja Grün received her diploma in physics from the Eberhard Karls University in Tübingen (1991) and her PhD in physics from Ruhr University Bochum (1996). After being a postdoc at the Hebrew University (Jerusalem) and at Max-Planck Institute in Frankfurt (M), she became a junior professor at Freie University Berlin in 2002, and unit/team leader at RIKEN Brain Science Institute (Tokyo) in 2006. Since 2011, she is a full professor at RWTH Aachen University and leads the group Statistical Neuroscience (INM-6, Research Center Jülich), and was appointed director of INM-6/INM-10 in 2018. Her work focuses on the development of analysis strategies and tools that uncover concerted activity in massively parallel electrophysiological recordings from the cortex, which led to the additional focus on research data management.

Thomas Wachtler
Department Biologie II, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany
wachtler@bio.lmu.de
https://orcid.org/0000-0003-2015-6590

Thomas Wachtler has a background in physics and received his diploma and doctoral degree from the University of Tübingen. He was a postdoctoral researcher at the Salk Institute for biological Studies and at the universities of Freiburg and Marburg. His research interests are in the neural mechanisms of sensory processing with focus on vision. In his research, he combines experimental and computational approaches, including electrophysiology, psychophysics, and computational modeling to study the neural principles of processing and coding in the visual system and how they relate to the properties of the sensory environment and to perceptual phenomena. He is also working on neuroinformatics developments in the context of the International Neuroinformatics Coordinating Facility. Since 2009, he has been the Scientific Director of the German Neuroinformatics Node at LMU Munich, leading developments of tools and services for research data management in neuroscience.

Hansjörg Scherberger
Neurobiology Laboratory, Deutsches Primatenzentrum GmbH, Kellnerweg 4, 37077 Göttingen, Germany
Department of Biology and Psychology, University of Goettingen, Goettingen, Germany
hscherb@gwdg.de
https://orcid.org/0000-0001-6593-2800

Hansjörg Scherberger heads the Neurobiology Laboratory at the German Primate Center and is a professor for primate neurobiology at Göttingen University (since 2008). He received his diploma in mathematics (1993) and his medical doctor degree (1996) from Freiburg University, Germany, and subsequently was trained in systems electrophysiology at the University of Zurich (1995–1998) and the California Institute of Technology (1998–2003) before leading a research group at the Institute of Neuroinformatics at Zurich University and ETH (2004–2009). His research focuses on neural coding and decoding of hand movements, their interactions with sensory systems, and he develops brain-machine interfaces to read out movement intentions for the development of neural prosthetics to restore hand function in paralyzed patients.